



Aberystwyth University

Constraint-based optimisation tools for semi-automated refinement of genome-scale yeast metabolic models

Pir, Pinar; Dobson, Paul D.; Smallbone, Kieran; Mendes, Pedro; King, Ross Donald; Lu, Chuan; Oliver, Stephen G.; Clare, Amanda

Publication date:

2010

Citation for published version (APA):

Pir, P., Dobson, P. D., Smallbone, K., Mendes, P., King, R. D., Lu, C., Oliver, S. G., & Clare, A. (2010).

Constraint-based optimisation tools for semi-automated refinement of genome-scale yeast metabolic models.

General rights

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400

email: is@aber.ac.uk

Constraint-based Optimisation Tools for Semi-automated Refinement of Genome-scale Yeast Metabolic Models

Chuan Lu¹ Pinar Pir², Kieran Smallbone³, Paul D. Dobson⁴, Amanda Clare¹,

Pedro Mendes³, Stephen G. Oliver², and Ross D. King¹,

¹Department of Computer Science, Aberystwyth University, UK, ²Cambridge Systems Biology Centre, Department of Biochemistry, University of Cambridge, UK,

³Manchester Centre for Integrative Systems Biology, University of Manchester, UK, ⁴School of Chemistry, University of Manchester, UK

Introduction

Motivation

- Genome-scale metabolic network models are useful for analysing the cellular behaviour of organisms
- Semi-automated procedure for model validation and refinement are important for quality assurance in such models
- Computational tools for iterative model validation and optimisation are necessary to assist hypothesis generation and evaluation

Genome scale metabolic models for *S. cerevisiae*

- A consensus reconstruction: Yeast1, community driven, rigorously evidenced, well annotated [1]
- Further development: Yeast4, expanded from Yeast1, with improved representation of metabolic transport, lipid metabolism, etc. [2]
- Yeast4: 1102 unique metabolite reactions, and 924 metabolites located in 15 cellular compartments

Applications

Computational tool implementation

- Implemented in Python, using CPLEX, glpk, IpSolve as LP/MILP solvers
- Read/write models in SBML format
 - Model stored in bipartite graph and/or stoichiometry matrices
 - Suitable for both FBA and logical model simulations
- Converting model network (bi-level) optimisation problem to constraint-based optimisation problem: LP, MILP.
- Algorithms for gap filling, OMNI
- Search algorithms for graph traverse, and identification of minimal models

Model validation with experimental data

- Single deletion data under minimal medium
- Wildtype growing under different conditions: RobotScientist's [3] automated titration experiments on yeast utilising amino acids as sole C/N source
- Awareness of data quality issue

Methods

Framework of Flux Balance Analysis (FBA)

- Identification of flux distribution using stoichiometry model, assuming steady states, with constraints on mass balance and thermodynamics to maximise/minimise an objective function (e.g. to max growth rate)
- Utilisation of constraint-based optimisation, linear/nonlinear programming (LP/NLP), mixed integer linear programming (MILP)

Gap filling

- Structural Gaps in metabolic networks
 - Reaction gaps, missing gene-protein-reaction associations, etc.
- Mechanisms to rescue reaction gaps
 - Reversibility; transport; biomass formation; metabolite exchange
 - Addition of missing reactions from reference model
- Identification of minimal set of reactions to add on, in order to restore biomass formation or blocked reactions [4]

Optimal Metabolic Network Identification (OMNI)

- Models under-constrained:
 - Reactions absent in yeast, irreversible or unfavorable under certain conditions, suppressed due to regulatory, etc.
- Bi-level constrained optimisation:
 - Minimisation of discrepancies between observations and predictions while maximising the growth rate [5]
 - Converting to MILP by exploiting duality for LP

Gap filling procedure

- Constraint-based optimisation and literature searching
- Solutions for 14 of 16 false inviable single deletions under minimal medium
- Further curation needed to fill in the missing reaction esp. for alternative pathways of ergosterol biosynthesis

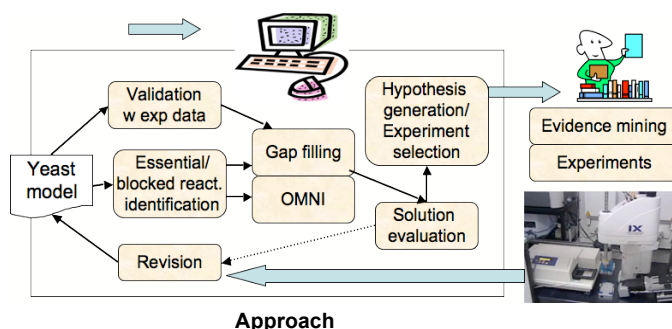
Single deletant	Non-producible biomass comp.	Revision suggestions	Revision field	Reference
IPT1, CSG2, PXA2, PXA1	MIP2C	Remove MIP2C from biomass, not essential for cell growth.	Biomass formation	PIMD: 9368028
SLC1	phosphatidylcholine triglyceride,...	1-acyl-sn-glycerol-3-phosphate acyltransferase: SLC1 => SLC4.	GPR	PIMD: 17726007
ILV6	L-valine, L-isoleucine	2-aceto-2-hydroxybutanoate synthase and acetolactate synthase: ILV6: ILV2 => ILV6: ILV2 JLV2.	GPR	PMID: 10213630
TGL2	triglyceride	Adding 2 putative transport reactions between cytosol and lipid particle.	Transport reaction	Gap filling algorithm
...

OMNI procedure

- More than 1 solution for 12 out of 48 false viable cases subject to OMNI
- Solution evaluation: *in-silico* simulation using phenotype data in SGD
- Application of a minimal set of revisions, resulting in:
 - True inviables increased by 12, at the cost of 1 extra false inviable
- Suggested revisions:
 - Constraining the reaction directionality
 - Removing reactions:
 - e.g. alternative pathway for quinolinate synthesis absent in yeast
 - Adding regulator rules to control reaction activation e.g. GALT and GALE activated only after sensing galactose
 - Testing in vivo by robot: auxotrophy experiments

Conclusions

- Proposed computational tools can **effectively search for (multiple) revision suggestions** for yeast metabolic models
- Semi-automated model refinement, supported with literature search and robot scientist experiments, helps to **improve the model in phenotype prediction**
- Future work**
 - Use of **logic programming** to integrate models with evidence from experimental data and constraint-based analysis
 - Learning GPR associations and regulatory rules and automated suggestion of experiments, either *in-silico* or *in-vivo***



Acknowledgement
The EU FP7 project of UNICELLSYS

Further information
Chuan Lu, PhD
Dept. of Computer Science
Aberystwyth University
Aberystwyth
SY23 3DB, Wales, UK
cul@aber.ac.uk

References

- Herrgård et al. A consensus yeast metabolic reconstruction obtained from a community approach to systems biology, Nature Biotechnol. 2008, 26
- Dobson et al. Further developments towards a genome-scale metabolic model of yeast. BMC Syst Biol, in press.
- King et al. The Automation of Science. Science 324(5923), 2009
- Reed et al. Systems approach to refining genome annotation. PNAS, 103:46, 2004.
- Herrgård et al. Identification of genome-scale metabolic network models using experimentally measured flux profiles. PLoS Comput Biol. 2006;2(7)